BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

ISSN 0036-8075

1 February 1985

Volume 227, No. 4686

The squabbles and the policy challenges will be more easily resolved if we understand their origin. In addition, we must focus our attention on the problem of institutional capacity and the health of capital resources. In comparison with what is available elsewhere, and what ought to be available to us, our environments are significantly worse then they were a quarter century ago. We owe to the next generation of students and faculty members an opportunity to do science as close to the forefront as all of us have been able to do it. Commitments only to the number of research grants next year, or to the total programmatic support of research in the federal budget, will not make that happen. It will only perpetuate the present liability, extend the divisions between researchers and institutions, and blunt the promise that our extraordinary way of doing science has created.

References and Notes

1. S. Bergstrom, opening address at the 1976 Nobel Festival, reprinted in *Inst. Med. Newsl.* (February 1977), pp. 3–4.
2. L. Feinberg, "Colleges bypass agencies to get federal funds," Washington *Post*, 5 June 1984, p. A-6.
3. House Committee on Appropriations, *Department of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Bill, 1983*, 97th Cong., 2nd sess., 29 September 1982, H. Rep. 894, p. 26.
4. "Statement of the Council on Federal Funding for Research Facilities and Instrumentation" (National Academy of Sciences, Washington, D.C., 30 October 1983).
5. *Fed. Am. Soc. Exp. Biol. Newsl.* **16**, 2 (March 1983).
6. Agreement between the General Hospital Corporations and Hoechst AG, 13 May 1981.
7. Association of American Universities, "The scientific instrumentation needs of research universities," a report to the National Science Foundation, June 1980.
8. The Carter Administration responded by requesting, as part of its fiscal year 1982 budget request, an additional $75-million appropriation to NSF for research equipment. This funding would have been only a small "down payment" on total liability estimated at over $1 billion ["Special analyses, budget of the United States Government, fiscal year 1982" (Office of Management and Budget, Washington, D.C., 1982), p. 314].
9. "Federal funds for research and development, fiscal years 1982, 1983, 1984" (NSF 83-319, Division of Science Resource Studies, National Science Foundation, Washington, D.C., 1983), vol. 32, p. 121.
10. B. R. Inman, *Aviat. Week Space Technol.* **116**, 10 (8 February 1982).
11. B. Schorr, "Breaking tradition, more colleges go directly to Congress for funds," *Wall Street Journal*, 5 March 1984, p. 29.
12. For a review of the rules and requirements of recovery of indirect costs, see Ad Hoc Committee on Government University Relationships in Support of Science, "Strengthening the government-university partnership in science" (National Academy Press, Washington, D.C., 1983), chapter 6, pp. 117–145.
13. The most recent such review was conducted by the San Francisco accounting firm of Peat, Marwick, and Mitchell in 1977.
14. As would be expected, the variance is markedly reduced when only similar institutions are considered. In addition to the important distinction between public and private, there are expected differences in rate between medical schools and other research areas, and between large- and small-volume performers. If one considers only those research universities that are (i) private, (ii) have medical schools and include them in the rate, and (iii) have large research contributions from both medical and nonmedical components, the following indirect cost rates may be compared for fiscal year 1984: University of Chicago, 69.0; Columbia University, 69.7; University of Pennsylvania, 65.0; Stanford University, 69.0; and Yale University, 68.0.
15. The underfunding strategy is disclosed in a letter from W. F. Raub, deputy director for extramural research and training at the National Institutes of Health, to A. Merritt, director of the Office of Research Administration at the University of Pennsylvania, in February 1984. Raub states, "While most Institutes are making only 1 to 2 percent reductions, the National Institute for Arthritis, Diabetes, and Digestive and Kidney Disease has found it necessary to make a larger reduction to fund its proportion of the approximately 5000 grants that the National Institutes of Health will be awarding in FY 1985."
16. These growth rate measurements were made from audited Stanford data on the actual indirect cost pools. Similar figures for the cost allocations would differ less, because in the process of arriving at the latter, each cost category is subject to proportional cross-allocations from the others. Thus, for example, general administration receives a cross-allocation from the operations and maintenance pool, so that it will include the costs of maintaining and heating space used for that activity. The effect of this cross-allocation will be to "load" the purely administrative costs with the more rapidly inflating building-related costs. The indirect cost pools themselves are uncontaminated by this effect.

RESEARCH ARTICLE

Nucleotide Sequence and Expression of an AIDS-Associated Retrovirus (ARV-2)

Ray Sanchez-Pescador, Michael D. Power, Philip J. Barr
Kathelyn S. Steimer, Michelle M. Stempien
Sheryl L. Brown-Shimer, Wendy W. Gee, Andre Renard
Anne Randolph, Jay A. Levy, Dino Dina, Paul A. Luciw

A wide variety of diseases in many animal species are a consequence of infection by retroviruses (*1*). A distinct group of human retroviruses has been isolated from patients with the acquired immune deficiency syndrome (AIDS) and individuals with related conditions, such as persistent lymphadenopathy. Several independent isolates, called lymphadenopathy-associated virus or LAV (*2*), human T-cell lymphotropic virus type III or HTLV-III (*3*), and AIDS-associated retrovirus or ARV (*4*) by the laboratories of origin, are similar with respect to morphology, cytopathology, requirements for optimum reverse transcriptase activity, at least some antigenic properties, and some restriction endonuclease cleavage sites in viral DNA. Epidemiological studies show that infection by one of these viruses may be a necessary condition for the development of AIDS, although predisposing factors may contribute to the onset of the disease (*3–10*).

Molecular clones of HTLV-III, LAV, and ARV-2 have been described (*11*, *12*). These clones provide material for analyses of viral structure, viral replication, and mechanisms of pathogenesis as well as for measurements of similarities and differences among the retroviruses associated with AIDS and with other retroviruses. In this report, the genetic structure of an ARV isolate is established from the sequences of molecular clones of ARV-2 DNA (*12*) and from the partial sequence of virion proteins.

The DNA sequence of ARV-2. Proviral DNA and circular unintegrated viral DNA species from ARV-2 infected cells have been cloned in bacteriophage λ (*12*), and the structures of five recombinant phage containing ARV-2 DNA were characterized (Fig. 1). The nucleotide sequence of various regions of each of these molecular clones was determined and used to establish the complete sequence of ARV-2 DNA. The sequence variations in ARV-2 DNA in these phage are presented in Table 1.

Long terminal repeat regions (LTR's). The LTR's of retroviruses participate in the integration of the virus with the host cell and in the regulation of transcription of viral genes (*13–15*). To define the

The authors are at Chiron Research Laboratories, Chiron Corporation, Emeryville, California 94608, except for Jay A. Levy who is at the Cancer Research Institute, University of California, School of Medicine, San Francisco, California 94143.

precise boundaries of the LTR sequences, we compared the junctions with host-cell DNA in the sequences of λ-9B, λ-7A, λ-8A, and λ-7D (Fig. 1). The LTR of ARV-2 is 636 bp and is bounded by an inverted repeat of 3 bp (CTG) (Fig. 2). The sizes of the inverted repeat at the ends of the LTR's of the other human retroviruses, HTLV-I and HTLV-II, are 2 bp (*16, 17*). Integration of proviruses did not occur in a specific site in the host cell genome since adjacent cell DNA sequences in λ-9B, λ-8A, and λ-7D were unique (data not shown). Preceding the rightward (3') LTR is a polypurine tract of 16 bp beginning at position 8632 (Fig. 2). Polypurine tracts are similarly positioned in other retroviruses and play an important role in the initiation of plus-strand DNA synthesis (*15*). Immediately downstream from the leftward (5') LTR is a sequence of 18 bp that is complementary to 18 bases of a transfer RNA–lysine ($tRNA^{lys}$) species (Fig. 2). Initiation of minus-strand DNA synthesis in retroviruses requires a host cell tRNA molecule as a primer (*15*). MMTV (mouse mammary tumor virus) also requires a $tRNA^{lys}$ molecule (*18*), whereas other known mammalian retroviruses including HTLV-I and HTLV-II have a tRNA-proline primer (*16, 17, 19*).

Contained within the LTR's of retroviruses are signals that control initiation and processing of viral transcripts (*13–15*). The cap site and a portion of the leader sequence are specified by the LTR. A primer-extension experiment in which we used purified virion RNA identified the 5'-end of ARV-2 RNA (Fig. 3). Thus, the ARV-2 LTR (R-U5 region) contributes 182 bp to the leader (Fig. 2). Many genes of eukaryotic cells and viruses contain a TATA box about 25 bp upstream from the start of the transcript (*20*); the TATA box is important for positioning the start site of transcription (*20, 21*). In the ARV-2 LTR sequence, a TATA box is located at −29 to −25. A 13-bp palindrome, at −25 to −13, overlaps the 3'-end of the ARV-2 TATA box; the significance of this structural feature is not known. Another common element of eukaryotic transcriptional units, a CAAT box, is usually positioned 60 to 70 bp upstream from the cap site (*22*). A similar feature is not present in the ARV-2 LTR.

A consensus sequence that signals addition of polyadenylated tails, AATAAA (*23*), is located in the rightward ARV-2 LTR at position 9174 to 9179 (Fig. 2). Further downstream in the LTR, between 9203 to 9224, is a region that is devoid of A residues (Fig. 2). The site of addition of polyadenylated [poly(A)] tails in the LTR's of many retroviruses is followed by a region of 20 to 30 bp that is also deficient in adenylic acid residues (*19*). For several eukaryotic genes and retroviruses, including MuLV (murine leukemia virus), MMTV, RSV (Rous sarcoma virus), and RAV-0 (Rous-associated virus), the dinucleotide CA is located at the poly(A) addition site (*19*). These comparisons were used to propose a tentative poly(A) addition site at positions 9198 in the rightward ARV-2 LTR (Fig. 2).

The enhancer element, generally located upstream from the TATA box, has been shown to be an important feature of transcriptional regulation for some eukaryotic genes and viruses (*24–28*). Large repeats, characteristic of some retroviral enhancers, are not present in the ARV-2 LTR. A close fit for the proposed consensus sequence for enhancer elements, (G) TGG^{AAA}_{TTT} (G) (*29*), is not found in the ARV-2 LTR.

The gag *gene.* The *gag* region of retroviruses encodes the internal structural proteins of the virion (*30*). A precursor polypeptide is synthesized and subsequently cleaved to yield mature *gag* proteins (*30*). The DNA sequence of ARV-2 predicts a *gag* precursor of 502 codons initiating at the ATG at position 337, the first ATG in the proposed full-length ARV-2 RNA (Fig. 2). To verify the use of this reading frame and to identify virion proteins as products of *gag*, we determined partial amino acid sequences of two virion proteins, p25 and p16, detected with serum from an AIDS patient (Fig. 4) but not with normal human

Table 1. Polymorphism of the λ recombinants shown in Fig. 1.

Position*	7A	7D	8A	8B	9B
−123	G		A		G
−115	G		A		G
3789			A	G	
4223			T	C	
5761	G	A			G

*Numbering system as described in Fig. 2.

Abstract. *The nucleotide sequence of molecular clones of DNA from a retrovirus, ARV-2, associated with the acquired immune deficiency syndrome (AIDS) was determined. Proviral DNA of ARV-2 (9737 base pairs) has long terminal repeat structures (636 base pairs) and long open reading frames encoding* gag *(506 codons),* pol *(1003 codons), and* env *(863 codons) genes. Two additional open reading frames were identified. Significant amino acid homology with several other retroviruses was noted in the predicted product of* gag *and* pol, *but ARV-2 was as closely related to murine and avian retroviruses as it was to human T-cell leukemia viruses (HTLV-I and HTLV-II). By means of an SV-40 vector in transfected simian cells, the cloned* gag *and* env *genes of ARV-2 were shown to express viral proteins.*

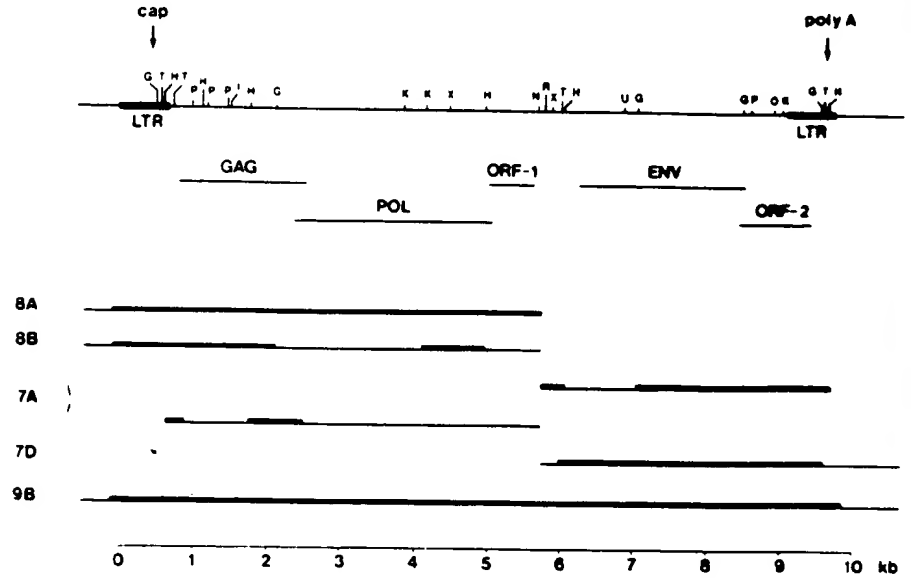
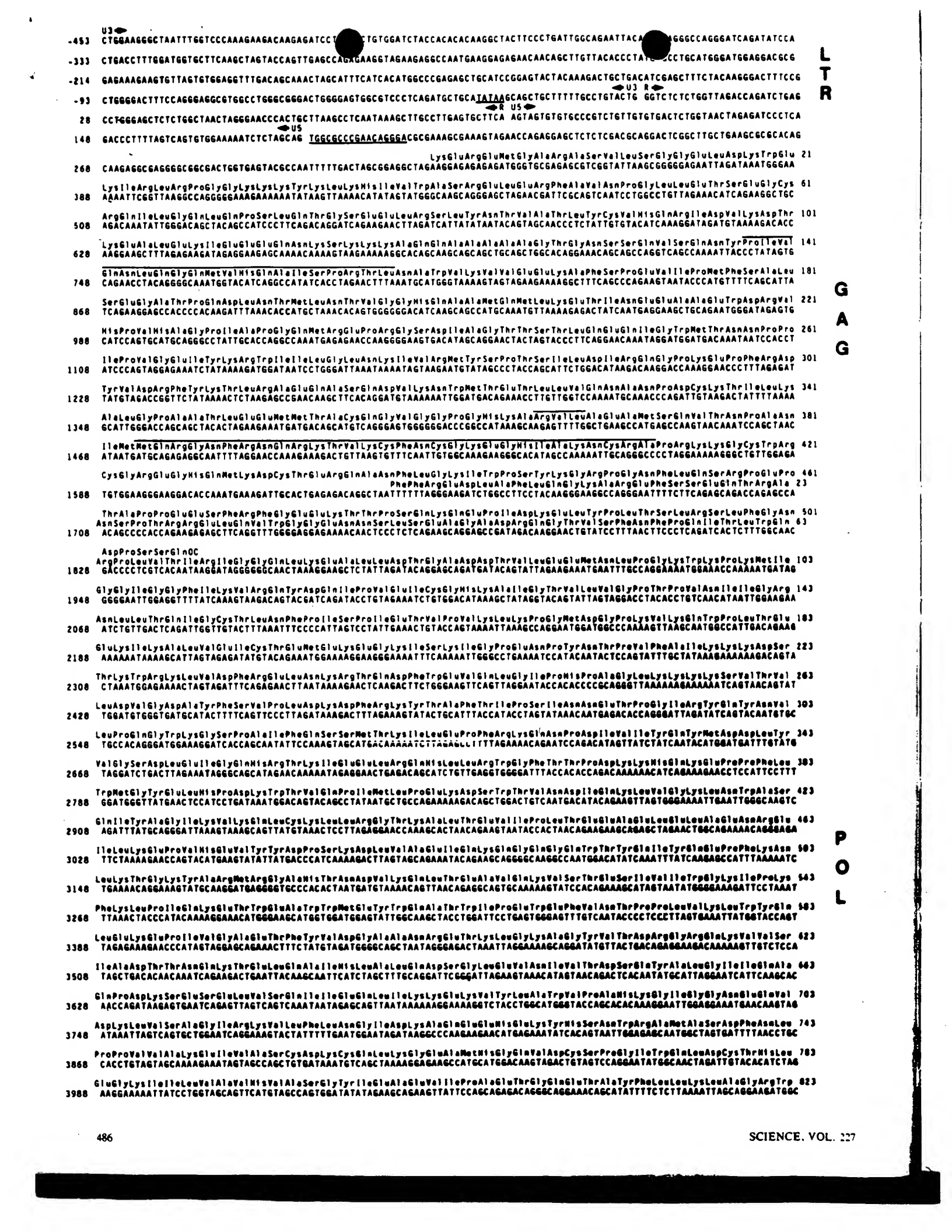


Fig. 1. Restriction endonuclease map of ARV-2. Five recombinant λ clones were isolated (*12*) and used to determine the nucleotide sequence of ARV. Clones 8A, 8B, 7D, and 9B represent integrated DNA. Clone 7A is from unintegrated DNA (*12*). The heavy lines indicate regions that were sequenced in each clone. The regions that encompass the *gag*, *pol*, and *env* ORF's as well as two additional open reading frames are indicated.


LTR
GAG
POL

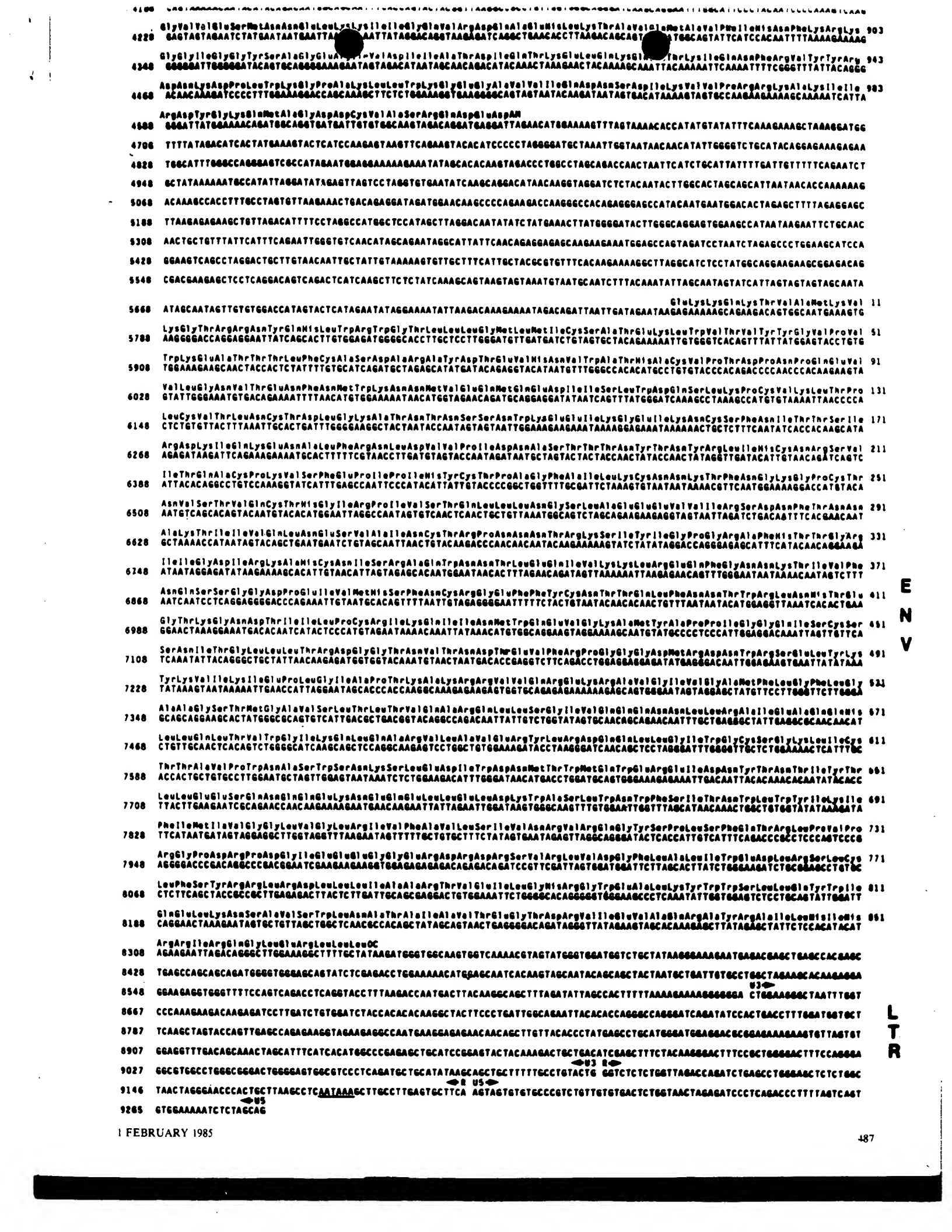

ENV
LTR

Fig. 2 (pages 486 and 487). Nucleotide sequence of ARV-2 DNA. The predicted amino acid sequences for the products of the *gag*, *pol*, and *env* genes are indicated. The U3, R, and U5 regions of the LTR's are also designated. The cap site, as determined from the experiment shown in Fig. 3, is position +1. A 3-bp inverted repeat at the ends of the LTR, the TATA box at position −29, the sequence complementary to the 3'-end of the $tRNA^{lys}$ at position 183, and the polyadenylation signal at position 9174 are underlined. The overlines indicate the amino acid sequences determined from virion proteins (Fig. 4). The nucleotides at the beginning of each line are numbered, and the amino acids at the end of each line are indicated. **Methods**: Restriction enzyme DNA fragments of recombinant phage DNA (Fig. 1) were isolated after electrophoresis in polyacrylamide or agarose gels, cloned into M13 vectors, and used as templates for DNA sequencing by the dideoxy chain termination method (*50*). Oligonucleotide primers for sequencing were chemically synthesized by solid-phase phosphoramidite chemistry on an Applied Biosystems 380A machine. The limits of the LTR's were established by comparing the sequence of both ends of proviral DNA as well as the sequence of a permuted clone (7A in Fig. 1). For protein sequencing, 0.38 mg of purified virus was subjected to electrophoresis on a 12 percent polyacrylamide Laemmli gel and the bands corresponding to p16*gag* and p25*gag* were cut out and electroeluted by the method of Hunkapiller *et al.* (*51*). NH_2-terminal microsequencing of these proteins was carried out as described by Hunkapiller *et al.* (*52*). COOH-terminal analysis was by the carboxypeptidase digestion procedures of Hayashi (*53*). The compiled ARV-2 DNA sequence, including both copies of the LTR, is 9737 bp in length. The analysis of the genetic organization of ARV-2 draws on comparisons with other retroviruses. For these comparisons we used computer programs such as MALIGN to identify homologous regions among DNA sequences and protein sequences. Structural relations were also investigated: predicted proteins from ARV-2 open reading frames were analyzed for hydropathy patterns by the method of Hopp and Woods (*54*) and for specific structural features by a modification of the method of Chou and Fasman (*55*). These two parameters were combined to determine regions of a protein that may be on the surface, particularly loops composed of hydrophilic residues.

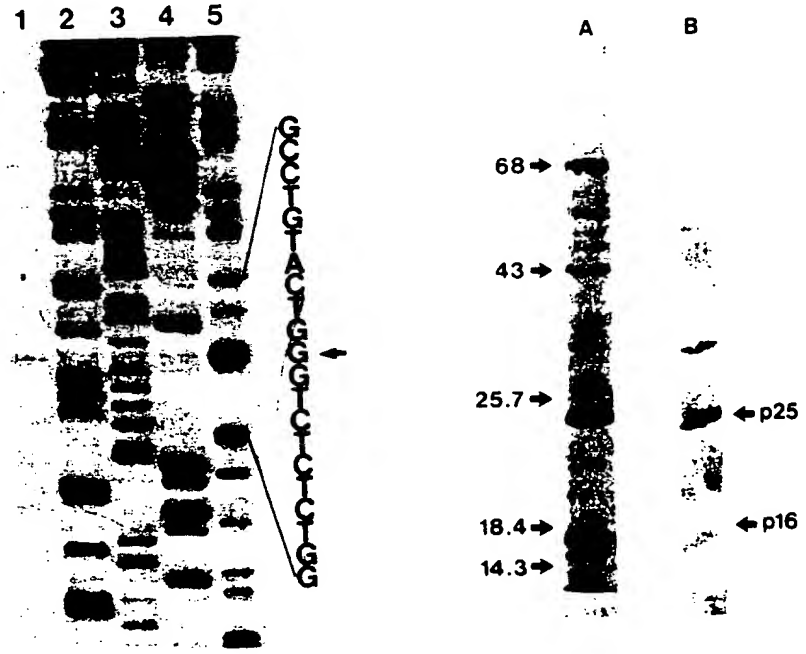


Fig. 3 (left). Identification of the 5'-end of ARV-2 RNA. Viral RNA was isolated from virions (*12*) and used as a template for Klenow fragment of DNA polymerase I with the synthetic oligonucleotide 5'GGGCACACACTACTTGAAGC as a primer. An M13 clone containing the leftward LTR of ARV-2 was also primed with the same oligonucleotide in the presence of dideoxynucleotides (*50*). Both reactions were resolved on a sequencing gel. Lane 1 corresponds to the primer extension reaction with ARV-2 RNA template. Lanes 2, 3, 4, and 5 correspond to C, T, A, and G, respectively, of the sequencing reactions of the M13 recombinant clone.
Fig. 4 (right). Polypeptides of purified virus. Gradient purified ARV-2 (5 μg per lane) was subjected to electrophoresis on a 12 percent polyacrylamide gel according to the method of Laemmli (*56*). Lane A, staining with Coomassie brilliant blue. Lane B (immunoblot), polypeptides transferred to nitrocellulose (*57*) and treated first with a 1:500 dilution of serum from an AIDS patient (EW5111 reference serum from P. Feorino, Centers for Disease Control, Atlanta, Georgia) and then with a 1:200 dilution of horseradish peroxidase–conjugated goat antiserum to human immunoglobulin G (Cappel Laboratories, No. 3201-0081). The color substrate was HRP Color Development Reagent (containing 4-chloro-1-napthol; Bio-Rad). The molecular weights of protein markers subjected to electrophoresis in parallel lanes are shown in kilodaltons on the left. P25 and p16 indicate the bands that correspond to p25*gag* and p16*gag* that were used as substrates for amino acid sequencing.

control serum (data not shown). Virion proteins were isolated from a polyacrylamide gel and the first 30 amino acids at the NH_2-terminus of p16 and the first 20 of p25 were determined by gas-phase microsequencing. Alignment with the DNA sequence (Fig. 2) suggests that the first *gag* polypeptide is 134 amino acids in length and may correspond to a p12*gag* virion protein species seen on polyacrylamide gels (unpublished results). The NH_2-terminus of p25 is generated by a cleavage between Tyr-138 and Pro-139 (Fig. 2). Proline is present at the NH_2-terminus of at least three other major retroviral *gag* proteins (p25*gag* of HTLV-I, p27*gag* of RSV, and p30*gag* of MuLV) (*16*, *19*). A protease with this cleavage specificity has not yet been identified in ARV-2, but this activity can be encoded by a retrovirus (*30*). The carboxyl terminus of p25*gag* was determined by digestion with carboxypeptidase and yielded the sequence Arg-Val-Leu (amino acids 367, 368, and 369, respectively). The NH_2-terminus of p16 is generated by cleavage between Met-383 and Met-384 (Fig. 2). Processing at this site may involve chymotrypsin or a chymotrypsin-like enzyme, which is believed to process part of the *gag* precursor polypeptide in other retroviruses (*30*). The COOH-terminus of p16 probably occurs at Gln-506 since a translational stop codon follows (Fig. 2), although further proteolytic processing could also be involved.

A small amount of amino acid sequence homology is noted when p25*gag* of ARV-2 is compared to p24*gag* of HTLV-I (*16*) (data not shown). This homology involves the position of two cysteine (C) residues relative to the COOH-terminal of both proteins (Fig. 2). Also, four of five amino acids at the COOH-terminus of p25*gag* of ARV-2 match those at the COOH-terminus of p24*gag* of HTLV-I (Fig. 2) (*16*). A preponderance of hydrophilic residues characterizes these proteins.

Sequence comparisons of p16*gag* of ARV-2 with p16*gag* of HTLV-I (*16*), p12*gag* of RSV (*19*), and p15*gag* of MuLV (*19*) reveal the best homology (Fig. 5). The relative positions of the five Cys residues in each of these three proteins are closely conserved and all three contain a high proportion of hydrophilic residues.

The pol *gene*. The *pol* region encodes the virion RNA-dependent DNA polymerase (reverse transcriptase). Several additional enzymatic functions related to replication are controlled by this region, including ribonuclease H, a DNA endo-

nuclease, and, in some retroviruses, protease (*15*, *30*). An open reading frame of 1003 codons appears to be the ARV-2 *pol* domain (Fig. 2). Some homology at the protein level is observed in the NH_2-terminal portions of the predicted *pol* genes of ARV-2, HTLV-I (*16*, *31*), RSV (*19*), and MuLV (*16*) (Fig. 6). This region is also homologous to portions of the putative viral polymerases of hepatitis B viruses and cauliflower mosaic virus (*31*). Analysis of the remainder of the *pol* genes of ARV-2, HTLV-I, RSV, and MuLV demonstrates appreciable homology in protein structure and sequence near the COOH-termini (*16*, *19*, *32*, *33*) (Fig. 7). A 32-kD polypeptide is produced by proteolytic processing near the COOH-terminus of the RSV *pol* polypeptide precursor (*33*). Alignments of shared amino acids in this region of the ARV-2 *pol* gene (in particular, Cys residues) with the defined NH_2-terminus of p32 of RSV (*33*) permits tentative identification of a processing site for the counterpart protein (Fig. 7).

The env *gene*. The *env* region encodes the major glycoprotein found in the membrane envelope of the virus and in the cytoplasmic membrane of infected cells (*30*). Retroviral *env* proteins arise generally from a precursor polypeptide that is processed at two or more sites: the first processing event removes a signal peptide of about 30 amino acids and the second yields a COOH-terminal polypeptide containing a hydrophobic stretch (about 22 amino acids) that spans the membrane and is followed by a hydrophilic cytoplasmic anchor (*30*). Results of transient expression experiments in mammalian cells (see Fig. 8) indicate that serologically reactive ARV-2 *env* protein is initiated downstream from the Sst I site at position 5555 to 5560 (Fig. 2). We propose that the ATG at position 5779 (*34*) initiates the *env* precursor, but direct determination of the NH_2-termini of the *env* precursor polypeptide and of processed forms will ultimately be required to establish the biogenesis of *env* proteins. Two other potential initiation codons are near the 5′-end of the same long open reading frame (863 codons) proposed to encode the ARV-2 *env* protein (positions 5845 and 5851, Fig. 2).

Secondary structure analysis shows that the COOH-terminal region is organized into predominantly α-helices and β-sheets; the NH_2-terminal half appears to have many hydrophilic loop regions (see legend to Fig. 2); similar structural properties characterize the domains of *env* gene products of other retroviruses (data not shown). A tentative assignment of a processing site for ARV-2 *env* includes the sequences Lys-Arg-Arg or Lys-Arg (Fig. 9). Which of these sites is used remains to be determined. Processing in this region will generate final products of 59 and 42 kD without accounting for carbohydrates residues. The NH_2-terminal and COOH-terminal portions contain, respectively, 26 and 5 potential NH_2-linked glycosylation sites (Asp-X-Thr, Asp-X-Ser) (Figs. 2 and 9). Cysteine residues are asymmetrically distributed as in other retroviral *env* gene products (*19*). The NH_2-terminal domain has

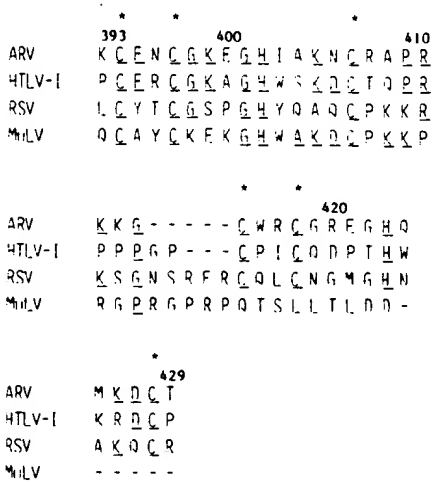


Fig. 5 (left). Homology of amino acids in regions of the *gag* gene of ARV-2, HTLV-I, RSV, and MuLV. Identical amino acid residues are underlined. Positions of cysteines are noted with asterisks. ARV-2: p16*gag*, amino acid 14 to 51 (Fig. 2). HTLV-I: p12*gag*, amino acid 12 to 50 (*10*). RSV: p12*gag*, amino acid 20 to 61 (*16*). MuLV: p10*gag*, amino acid 25 to 60 (*16*). Numbers indicate amino acid positions (Fig. 2).

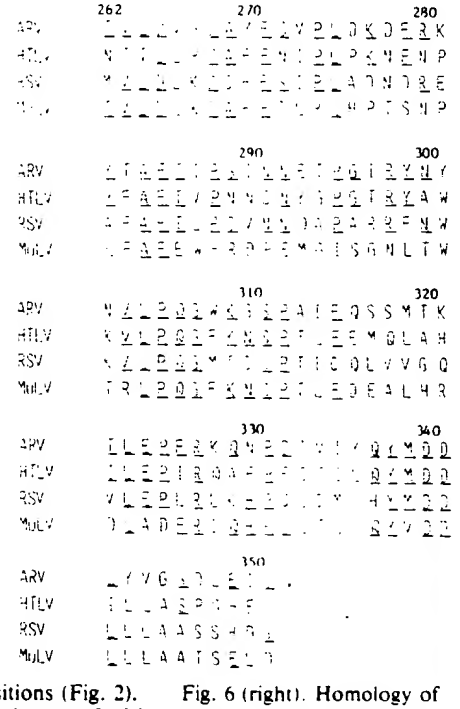
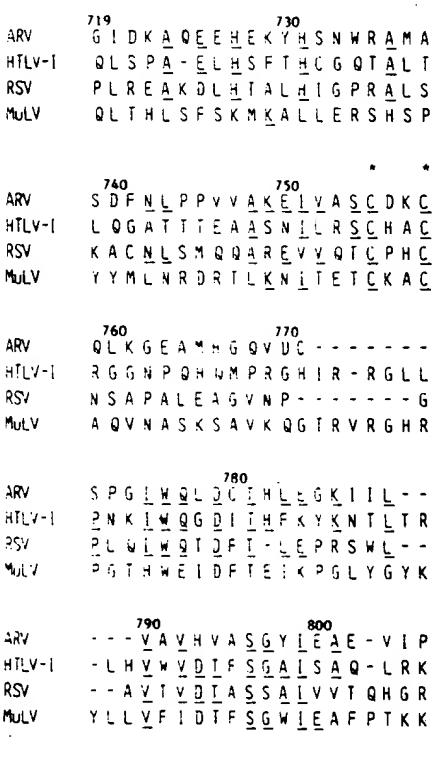


Fig. 6 (right). Homology of amino acids in the NH_2-terminal portion of the *pol* genes of ARV-2, HTLV-I, RSV, and MuLV. Identical amino acid residues are underlined. ARV-2: amino acid 262 to 352 (Fig. 2). HTLV-I: amino acid 110 to 196 (*10*). RSV: amino acid 113 to 177 (*16*). MuLV: amino acid 265 to 351 (*16*). Numbers indicate amino acid positions (Fig. 2).



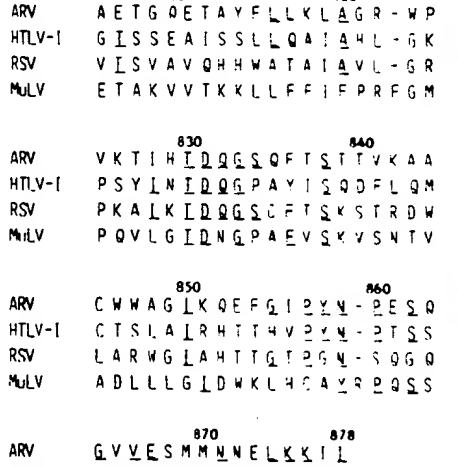


Fig. 7. Homology of amino acids in the COOH-terminal portion of the *pol* genes of ARV-2, HTLV-I, RSV, and MuLV. Identical amino acid residues are underlined. Positions of cysteines are noted with asterisks. ARV-2: 719 to 878 (Fig. 2). HTLV-I: amino acid 599 to 766 (*10*). RSV: amino acid 568 to 743 (*16*). MuLV: amino acid 846 to 1019 (*16*). Numbers indicate amino acid positions (Fig. 2).

Table 2. Summary of homologies of ARV-2 with other retroviruses. Homologies are given as percentages from the MALIGN program.

Virus	ARV-2 *gag* (amino acid 393 to 429, Fig. 5)		ARV-2 *pol* (amino acid 303 to 390, Fig. 6)		ARV-2 *pol* (amino acid 719 to 878, Fig. 7)	
	Amino acid	Nucleotide	Amino acid	Nucleotide	Amino acid	Nucleotide
HTLV-I	46	7	42	11	28	19
RSV	39	7	49	15	28	12
MuLV	27	10	35	15	12	23

18 Cys residues and the COOH-terminal portion has 3 Cys residues. Two large hydrophobic regions are evident in the COOH-terminal domain (Fig. 9). The rightward hydrophobic stretch is long enough (23 amino acids) to span membranes.

Expression of cloned ARV genes. In an attempt to obtain ARV antigens without the production of infectious virus, an SV40 vector system was used to express the candidate *gag* and *env* genes in transfected mammalian cells. The criterion for expression was serological reactivity of fixed cells with serum from AIDS patients in immunofluorescence tests. Recombinant SV40 plasmids containing these genes were transfected into 5×10^4 COS-7 monkey cells growing on microscope slides (Fig. 8); after 60 hours, cell monolayers were fixed and treated with AIDS patients' sera or normal human control sera and then with fluorescein-labeled goat antiserum to human immunoglobulin G (Fig. 8). Approximately 5 percent of cells transfected with pSV7c/gag showed a speckled pattern of immunofluorescence throughout the cytoplasm with AIDS patient serum EW5111 (Fig. 8A). Antiserum MC from a patient in the early stage of AIDS appeared not to react with cells transfected with pSV7c/gag (data not shown). By immunoblot analysis with proteins from purified ARV-2, antiserum MC was shown to have very low levels of antibody to p25*gag*, whereas antiserum EW5111 readily reacted to p25*gag*. Serum from normal individuals gave no appreciable fluorescence (data not shown) in cells transfected with pSV7c/gag cells transfected with the vector plasmid pSV7c, containing no ARV-2 DNA, did not fluoresce with any serum sample from AIDS patients. About 5 percent of cells (from 5×10^4 cells per microscope slide) transfected with pSV7c/env and treated with either EW5111 or MC antiserum showed bright immunofluorescence largely confined to the cytoplasm in a netlike pattern (Fig. 8B). These patterns may be a consequence of the fixation procedure or may indicate that viral *env* protein is localized in structures such as endoplasmic reticulum inside the cell. No fluorescence was observed in cells transfected with pSV7c/env and treated with normal human control sera (data not shown).

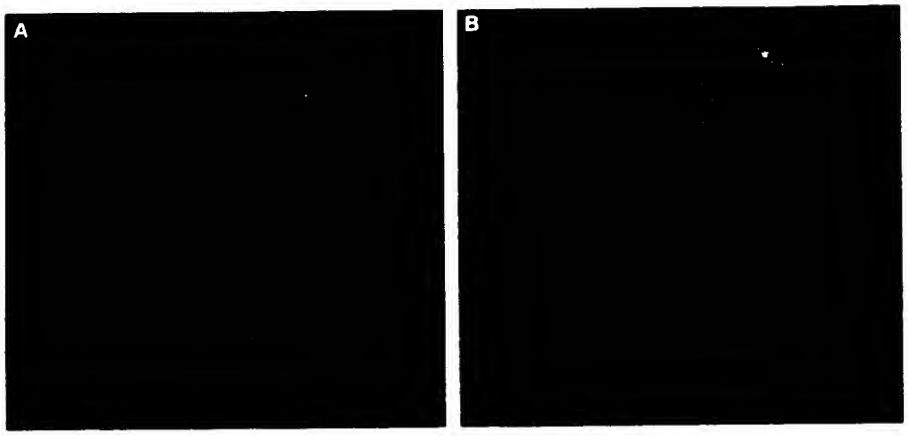


Fig. 8. Expression of cloned ARV genes in mammalian cells. ARV-2 DNA fragments containing the *gag* and *env* genes were prepared as follows: λ-7A DNA (Fig. 1) was digested with Sst I and Kpn I and the 3.1-kb *gag* DNA fragment was purified by electrophoresis in low-melting agarose gels (*7*); λ-7D DNA (Fig. 1) was digested with Sst I and Kpn I and the 3.2-kb *env* DNA fragment was similarly purified. Each of these fragments was cloned into a modified form of a plasmid containing the SV40 origin of DNA synthesis and the promoter and poly(A) addition regions of the SV40 early gene (*58*, *59*). Both ARV *gag* and *env* DNA fragments contain ATG start codons. pSV7c/gag utilized a TAA stop codon in SV40 DNA. pSV7c env has the TAA stop codon at the end of the open reading frame for *env* (Fig. 2). COS-7 monkey cells, expressing the SV40 early gene, were grown on glass microscope slides, transfected with plasmid DNA by the calcium phosphate coprecipitation method (*60*), incubated for 60 hours, and fixed in cold acetone. The fixed cell monolayers were treated for 1 hour at 37°C with a 1:200 dilution (in PBS with 5 percent fetal calf serum) of an AIDS reference serum (Fig. 4) or with a similar dilution of normal human control serum. Cells were washed in PBS and treated for 1 hour at 37°C with fluorescein-labeled goat antiserum to human immunoglobulin G (Cappel Laboratories). In all cases, sera were preadsorbed on normal COS-7 cells that had been fixed with 0.2 percent paraformaldehyde. Shown here are fluorescence photomicrographs (×630) of cells transfected with (A) pSV 7C/gag and (B) pSV7C/env. About 5 percent of cells in a monolayer expressed viral antigens.

Discussion. The complete DNA sequence of ARV-2 reveals a fundamental genetic structure similar to that of other retroviruses. Several features of ARV-2 indicate that it is no more closely related to the other human retroviruses HTLV-I and HTLV-II than it is to avian or murine retroviruses.

ARV-2 has an inverted 3 bp repeat (CTG . . . CAG) at the ends of the LTR. All other retrovirus LTR's have TG . . . CA at their ends as part of a 2- to 16-bp inverted repeat (*14*). The MuLV LTR has two direct repeats 72 bp long located in an internal position within the LTR (*19*). HTLV-II has several direct repeats, one of which is 21 bp long and is very similar to a 21-bp repeat in HTLV-I (*16*, *17*). RSV, however, is like ARV-2 and has no large direct repeats in its LTR (*19*). In the ARV-2 LTR, the proposed poly(A) addition site is 20 bp downstream from the consensus poly(A) addition signal, AATAAA (Fig. 2); thus, the R region is 97 bp long [measured from the cap site to the poly(A) site]. The poly(A) addition sites of MuLV and RSV are about 20 bp downstream from AATAAA found in each LTR; these viruses have R regions 68 bp and 21 bp, respectively (*19*). In contrast, in HTLV-I and HTLV-II, the AATAAA sequence is located upstream from the TATA box; R is 229 bp in HTLV-I and 287 bp in HTLV-II (*16*, *17*). ARV-2 and MMTV have a $tRNA^{lys}$ for priming minus-strand DNA synthesis (Fig. 2) (*18*); avian retroviruses use $tRNA^{trp}$ and other mammalian retroviruses use $tRNA^{pro}$ (*19*).

The *gag* regions of MuLV and RSV encode precursor polypeptides that are cleaved into at least four and five proteins, respectively (*30*). Both ARV-2 and HTLV-I encode a *gag* precursor that appears to give rise to three proteins (Fig. 2) (*16*). A small amount of homology of amino acid sequences was noted in the COOH-terminal portion of *gag* in these viruses; ARV-2, HTLV-I, and

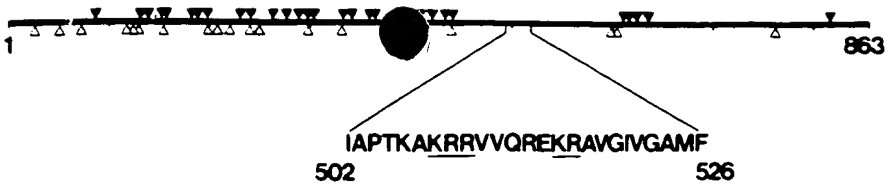


Fig. 9. Schematic diagram of ARV-2 *env* open reading frame. Numbers refer to amino acids in the open reading frame proposed for *env* (nucleotides 5755 to 8346, Fig. 2). Symbols: △, cysteine residues; ▼, potential *N*-glycosylation sites; ◆, hydrophobic regions. The two putative processing sites for generating NH_2- and COOH-terminal domains are underlined.

RSV were found to be similarly related in this assessment (Fig. 5 and Table 2).

Different retroviruses use different mechanisms to synthesize and translate the *pol* gene messenger RNA (*16*). Elucidation of *pol* biogenesis in ARV-2 will require detailed analyses of splicing patterns of viral mRNA in infected cells together with studies of the polypeptide intermediates. ARV may be different from all other retroviruses since the COOH-terminal end of the proposed *pol* gene does not overlap the NH_2-terminal end of the proposed *env* gene.

The predicted ARV-2 *env* polypeptide, like that of other retroviruses, has a hydrophilic NH_2-terminal domain and a COOH-terminal portion characterized by a long stretch of hydrophobic amino acids (23 amino acids long) (Fig. 9). The NH_2-terminal domain of ARV-2 *env* contains 26 potential glycosylation sites, an unusually high number when compared to other retroviruses: HTLV-I has 5 (*17*), HTLV-II has 6 (*35*), RSV has 17 (*24*), and MuLV has 7 (*24*). The extent and function of glycosylation in retroviral *env* proteins remain to be investigated.

In ARV-2 there are two additional open reading frames designated ORF-1 and ORF-2 (Fig. 10). Near the 5′-end of each open reading frame is an ATG that is flanked by purine residues at −3 and +4; thus, these ATG codons are potential start codons (*31*). HTLV-I (*16*), HTLV-II (*17*), and BLV (*36*) contain open reading frames that initiate beyond *env* and extend into the rightward LTR; this location is analogous to that of ORF-2 in ARV-2. Comparisons of ORF-2 in ARV-2 with counterpart regions in these other retroviruses revealed no apparent homology at the DNA and protein levels (data not shown). For HTLV-I and HTLV-II, these regions are expressed as proteins that are implicated in viral pathogenesis (*37*, *38*). Assessments of patterns of transcription and polypeptide synthesis will be essential to determine whether or not these ARV-2 open reading frames are expressed.

Certain taxonomic issues need to be addressed with respect to the relationships among the human retroviruses at the nucleotide sequence level. A probe

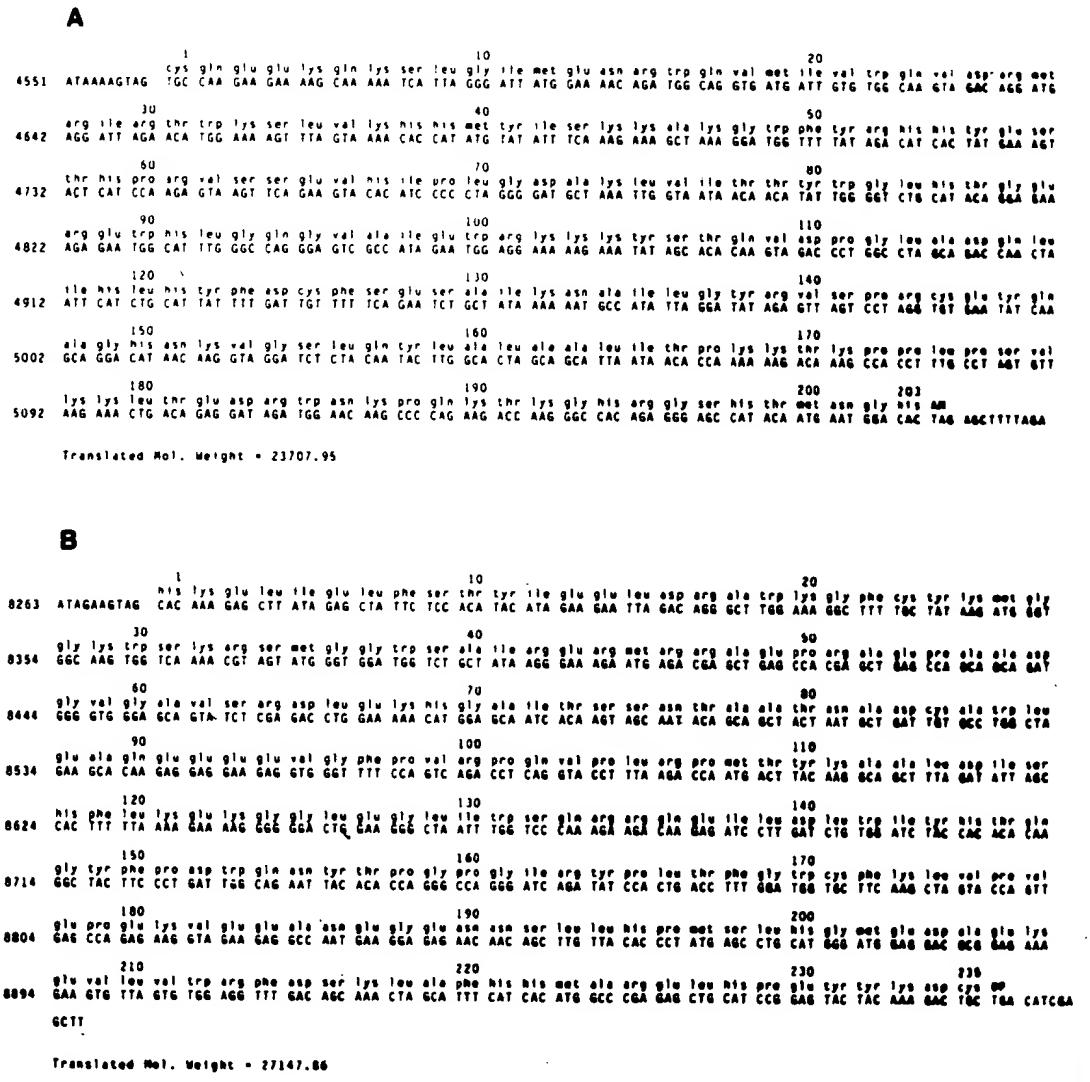


Fig. 10. Amino acid and DNA sequence of (A) open reading frame 1 (ORF-1) and (B) open reading frame 2 (ORF-2). The molecular weights are given in daltons. Nucleotides are numbered according to Fig. 2.

representative of ARV-2 anneals under high stringency conditions to restriction enzyme DNA fragments from cells infected with LAV or with HTLV-III (*39*). Thus, these three retroviruses are closely related. In addition, we have shown that the probe to ARV-2 anneals under high stringency conditions to proviral DNA of two independent isolates, ARV-3 and ARV-4 (*12*). At the protein level, very low homology is evident when ARV-2 genes are compared with those of HTLV-I (Figs. 5 to 7 and Table 2): homology at the nucleotide level is even lower because of degeneracy of codons (Table 2). In our assessments, ARV-2 appears to be no more closely related to these other human retroviruses than it is to RSV (Table 2). Subhuman primate endogenous viral sequences (*40*, *41*) are also distantly related to ARV *pol* (data not shown). Hybridization and annealing studies under very low stringency conditions demonstrated detectable homology of HTLV-III with HTLV-I and HTLV-II (*11*, *42*). Our homology assessments at the nucleotide level (Table 2) indicate that stable hybrids or duplexes cannot be formed between ARV-2 DNA and HTLV-I DNA under these conditions. These issues could be fully resolved by comparing the DNA sequences of the genomes of retroviruses associated with AIDS (LAV, HTLV-III, and ARV).

The pathology that attends ARV infection is a unique aspect of this retrovirus. Selective tropism for human T-helper cells, syncytia formation, and cell killing are characteristics of ARV infection in tissue culture cells (*2–4*, *43*). Attachment of virus to cell receptors and fusion of membranes are two properties controlled by the *env* gene that probably play a fundamental role in viral pathogenesis. The predicted sequence of ARV-2 *env* will be used to design mutagenesis experiments aimed at determining the function of *env* in attachment and fusion. LTR's of some avian and mammalian retroviruses have been shown to control tissue tropism, leukemogenicity, and specific disease patterns (*44–48*). Whether or not the ARV LTR plays a role in any of the pathologic manifestations associated with ARV infection remains to be established.

Sequence variations in ARV may be an important feature of viral pathogenesis that would enable the virus to evade host immune responses. Many viruses show sequence variation during passage. Infection of an animal with equine infectious anemia virus (EIAV) leads to differences in the *env* protein of progeny virus, probably as a consequence of immunological selective pressures in the host (*49*). Our studies of ARV have demonstrated sequence differences (i) in separate molecular clones of one ARV-2 isolate (Table 1) and (ii) in independent ARV isolates (*12*). Biological activity of cloned ARV-2 DNA has not yet been assessed by transfection of permissive cells. The generation of sequence variation in the ARV-2 genome can be studied by analyzing viruses recovered from different molecularly cloned ARV-2 DNA's. These approaches could provide insight into methods by which the viral infection could be prevented, modified, or eliminated.

References and Notes

1. N. Teich, J. Wyke, T. Mark, A. Bernstein, W. Hardy, in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y., 1982), p. 785.
2. F. Barré-Sinoussi *et al.*, *Science* **220**, 868 (1983).
3. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *ibid.* **224**, 497 (1984); R. C. Gallo *et al.*, *ibid.*, p. 500.
4. J. A. Levy *et al.*, *ibid.* **225**, 840 (1984).
5. J. A. Levy and J. Ziegler, *Lancet* **1983-II**, 78 (1983).
6. J. Schüpbach *et al.*, *Science* **224**, 503 (1984).
7. M. G. Sarngadharan, M. Popovic, L. Bruch, J. Schupbach, R. C. Gallo, *ibid.*, p. 506.
8. J. Laurence *et al.*, *N. Engl. J. Med.* **311**, 1269 (1984).
9. J. J. Goedert *et al.*, *Lancet* **1984-I**, 711 (1984).
10. F. Brun-Vezinet *et al.*, *ibid.* **1983-I**, 1253 (1983).
11. B. H. Hahn *et al.*, *Nature (London)* **312**, 167 (1984); M. Alizon *et al.*, *ibid.*, p. 757.
12. P. A. Luciw, S. J. Potter, K. Steimer, D. Dina, J. A. Levy, *ibid.*, p. 760.
13. H. M. Temin, *Cell* **27**, 1 (1981).
14. ———, *ibid.* **28**, 3 (1982).
15. H. E. Varmus and R. Swanstrom, in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), p. 369.
16. M. Seiki, S. Hattori, Y. Hirayama, M. Yoshida, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3618 (1983).
17. J. Sodroski *et al.*, *ibid.* **81**, 4617 (1984).
18. G. G. Peters and C. Glover, *J. Virol.* **35**, 31 (1980).
19. Appendix to *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), p. 1321.
20. E. B. Ziff and R. M. Evans, *Cell* **15**, 1463 (1978).
21. T. Yamamoto, D. deCrombrugghe, I. Pastan, *ibid.* **22**, 787 (1980).
22. A. Efstratiadis *et al.*, *ibid.* **21**, 653 (1980).
23. N. J. Proudfoot and G. G. Brownlee, *Nature (London)* **252**, 359 (1974).
24. [illegible] (1981).
25. P. G[illegible], R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 943 (1981).
26. M. R[illegible]er and M. Botchan, *Mol. Cell. Biol.* **3**, 325 (1983).
27. L. A. Laimins, G. Khoury, C. Gorman, B. Howard, P. Gruss, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6453 (1982).
28. P. A. Luciw, J. M. Bishop, H. E. Varmus, M. R. Capecchi, *Cell* **33**, 705 (1983).
29. H. Weiher, M. König, P. Gruss, *Science* **219**, 626 (1983).
30. C. Dickson, R. Eisenman, H. Fan, E. Hunter, N. Teich, in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, R. Weiss *et al.*, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), p. 513.
31. H. Toh, H. Hayashida, T. Miyata, *Nature (London)* **305**, 827 (1983).
32. I-M. Chiu, R. Callahan, S. R. Tronick, J. Schlom, S. A. Aaronson, *Science* **223**, 364 (1984).
33. D. P. Grandgenett, R. J. Knaus, P. J. Hippenmeyer, *Virology* **130**, 257 (1983).
34. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1983).
35. J. Sodroski, R. Patarca, D. Perkins, D. Briggs, T.-H. Lee, M. Essex, J. Coligan, F. Wong-Staal, R. C. Gallo, W. A. Haseltine, *Science* **225**, 421 (1984).
36. N. R. Rice *et al.*, *Virology* **138**, 82 (1984).
37. D. J. Slamon, K. Shimotohno, M. J. Cline, D. W. Golde, I.S.Y. Chen, *Science* **226**, 61 (1984).
38. T. H. Lee *et al.*, *ibid.*, p. 57.
39. M. Bryant and M. Gardner, personal communication.
40. T. I. Bonner, C. O'Connell, M. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4709 (1982).
41. T. A. Tamura, *J. Virol.* **47**, 140 (1983).
42. S. K. Arya *et al.*, *Science* **225**, 927 (1984).
43. D. Klatzmann *et al.*, *ibid.*, p. 59.
44. P. A. Chatis, C. A. Holland, J. W. Hartley, W. P. Rowe, N. Hopkins, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4408 (1983).
45. L. DesGroseillers, E. Rassart, P. Jolicoeur, *ibid.*, p. 4203; J. Lenz and W. A. Haseltine, *J. Virol.* **47**, 317 (1983).
46. P. N. Tsichlis *et al.*, *Mol. Cell Biol.* **2**, 1331 (1982).
47. W. A. Haseltine *et al.*, *Science* **225**, 419 (1984).
48. S. Y. Chen, J. McLaughlin, D. W. Golde, *Nature (London)* **309**, 276 (1984).
49. R. Montelaro, B. Parekh, C. Issel, A. Orrego, *J. Biol. Chem.* **259**, 10539 (1984).
50. J. Messing and J. Viera, *Gene* **19**, 269 (1982).
51. M. Hunkapiller, E. Lujar, F. Ostrander, L. Hood, *Methods Enzymol.* **91**, 227 (1983).
52. M. Hunkapiller, R. Hewick, W. Dreyer, L. Hood, *ibid.*, p. 399.
53. R. Hayashi, *ibid.* **47**, 84 (1977).
54. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824 (1981).
55. P. Y. Chou, and G. D. Fasman, *Annu. Rev. Biochem.* **47**, 251 (1978).
56. U. Laemmli, *Nature (London)* **227**, 680 (1970).
57. H. Towbin, T. Staehelin, J. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350 (1979).
58. M. Lusky and M. Botchan, *Nature (London)* **293**, 79 (1981).
59. O. Laub *et al.*, *J. Virol.* **48**, 271 (1983).
60. F. L. Graham and A. J. van de Eb, *Virology* **52**, 456 (1973).

61. We thank R. Blacher and F. Masuarz for viral protein purification and amino acids analysis, R. Najarian for help in the DNA sequencing, M. Powers for immunoblots, D. Parkes for restriction enzyme analysis, C. T. Lee-Ng for synthetic DNA purification, M. A. Wormstead for assistance in tissue culture, and P. Montes and D. Topping for preparation of the manuscript. We also thank L. Overby and our colleagues at Chiron for comments and support. H. E. Varmus (University of California, San Francisco) is acknowledged for helpful discussions. A. Renard is a visiting scholar from Laboratoire de Genie Gentique, Universite de Liège, 4000 Liège, Belgium.

5 December 1984; accepted 31 December 1984